

# Multivariate quadrature for representing cloud condensation nuclei activity of aerosol populations

Laura Fierce<sup>1,2</sup> and Robert L. McGraw<sup>1</sup>

<sup>1</sup>Climate and Environmental Sciences Department, Brookhaven National Laboratory

<sup>2</sup>Visiting Scientist Programs, University Corporation for Atmospheric Research

## Key Points:

- Multivariate quadrature resolves key features of multi-component aerosol populations
- Constrained maximum entropy distribution are used to reconstruct critical supersaturation distributions
- The combined technique reproduces CCN spectrum from particle-resolved model with high accuracy

arXiv:1612.06422v1 [physics.ao-ph] 19 Dec 2016

---

Corresponding author: Laura Fierce, [lfierce@bnl.gov](mailto:lfierce@bnl.gov)

## Abstract

Sparse representations of atmospheric aerosols are needed for efficient regional- and global-scale chemical transport models. Here we introduce a new framework for representing aerosol distributions, based on the quadrature method of moments. Given a set of moment constraints, we show how linear programming, combined with an entropy-inspired cost function, can be used to construct optimized quadrature representations of aerosol distributions. The sparse representations derived from this approach accurately reproduce cloud condensation nuclei (CCN) activity for realistically complex distributions simulated by a particle-resolved model. Additionally, the linear programming techniques described in this study can be used to bound key aerosol properties, such as the number concentration of CCN. Unlike the commonly used sparse representations, such as modal and sectional schemes, the maximum-entropy approach described here is not constrained to pre-determined size bins or assumed distribution shapes. This study is a first step toward a particle-based aerosol scheme that will track multivariate aerosol distributions with sufficient computational efficiency for large-scale simulations.

## 1 Introduction

Atmospheric aerosols remain a large source of uncertainty in prediction of changes and variability in the climate system [Myhre *et al.*, 2001; Kaufman *et al.*, 2002; Carslaw *et al.*, 2013; Stocker *et al.*, 2013]. Particle characteristics relevant for climate, such as optical cross sections and the critical supersaturation at which particles serve as cloud condensation nuclei (CCN), depend on particle size, shape, and chemical composition [e.g. Jacobson, 2001; Chung *et al.*, 2011; Cubison *et al.*, 2008; Ervens *et al.*, 2010]. Observations show tremendous variability in particle physical and chemical properties [e.g. Krieger *et al.*, 2012; Zhang *et al.*, 2014], but complex distributions of multicomponent particles are not easily represented in large-scale models. In this study, we seek a sparse representation of aerosol populations that resolves enough information about particle size and composition distributions to adequately represent climate-relevant properties, while resolving as little information as is necessary in order to minimize computational costs.

Many numerical models have been developed to simulate the aerosol evolution [e.g. Wexler *et al.*, 1994; McGraw, 1997; Jacobson, 1997; Penner *et al.*, 1998; Binkowski and Roselle, 2003; Stier *et al.*, 2005; Bauer *et al.*, 2008; Riemer *et al.*, 2009]. The simplest models track particle mass only, without any resolution of particle size or composition [e.g. Haywood *et al.*, 1997; Myhre *et al.*, 1998; Penner *et al.*, 1998; Cooke *et al.*, 1999; Lesins *et al.*, 2002]. At the other extreme, the Particle Monte Carlo Model (PartMC) [Riemer *et al.*, 2009] coupled to the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) [Zaveri *et al.*, 2008] tracks the evolution of thousands of individual particles, but is computationally expensive. Many large-scale models now include aerosol microphysical schemes that simulate the evolution of the particle size distribution, but resolve limited information about particle composition. Sectional models track the evolution of particles in different size bins, typically assuming that particles of the same size have identical composition [e.g. Wexler *et al.*, 1994; Jacobson, 1997]. Sectional models have also been extended to resolve multivariate distributions with respect to particle composition, such as particle hygroscopicity parameter or BC mass fraction, in addition to particle size [Oshima *et al.*, 2009; Matsui *et al.*, 2013; Zhu *et al.*, 2015], or to track separate aerosol populations [Kleeman and Cass, 2001; Jacobson, 2002]. Because sectional schemes simulate the evolution of a discretized representation of the full particle size distribution or, in the multivariate case, the particle size-composition distribution, they require tracking a large number of bins to resolve aerosol populations.

Moment-based models are a class of methods for simulating the evolution of probability density functions in which integral quantities over the distribution, such as radial moments, are tracked, rather than evolving the distribution itself [Hulburt and Katz, 1964; McGraw, 1997]. In this way, moment-based models represent the aerosol using an intermediate level of detail between the simple mass-based aerosol schemes and the more detailed particle-resolved and sectional representations. To date, moment-based aerosol representations have been implemented in global models as modal or monodisperse schemes that track only two low-order radial moments for each population [e.g. Stier *et al.*, 2005; Bauer *et al.*, 2008], typically aerosol number and volume or mass. Monodisperse aerosol models follow a similar approach [Pirjola *et al.*, 2003]. Modal models begin by assuming a specific

distribution shape, typically lognormal, track two moments per mode, and assume a prescribed geometric standard deviation. Particles within a mode are further assumed to contain identical mass fractions of the constituent aerosol species. However, comparison between modal and sectional aerosol schemes show that modal models can yield inaccurate representation of the particle size distribution [Mann *et al.*, 2012]. Further, inadequate resolution of variability in composition within each mode leads to errors in prediction of climate-relevant properties [Fierce *et al.*, 2016a,b].

The present study introduces a new framework for constructing aerosol representations in moment-based models, including modal models. Focusing on particle CCN activation properties, traditionally a difficult case for moment-based representations [Wright *et al.*, 2001, 2002], we show how efficient quadrature representations can be constructed from moments of aerosol distributions bivariate with respect to dry diameter and hygroscopicity parameter. By applying a transformation of variables we construct sparse quadrature point distributions of the aerosol in terms of the critical saturation for cloud droplet activation. Continuous distributions are obtained off line using the maximum entropy spectral representation constrained by the same moment/quadrature parameters that defined the sparse distribution. We show that this approach yields accurate prediction of CCN activity using only a limited number of moments.

## 2 Constructing sparse and continuous representations from distribution moments

Here we describe a new method for utilizing moments and moment-based, multivariate quadrature approximation to obtain sparse representations of benchmark PartMC-generated aerosol populations. Full size distributions, derived in terms of the sparse quadrature representation, are obtained using the principle of maximum entropy. Putting these two steps together we demonstrate that multivariate quadratures can be used to accurately compute CCN activities for the benchmark population.

### 2.1 Computing CCN activity for benchmark populations

The particle-resolved model PartMC-MOSAIC was used to generate realistically complex particle populations (see Appendix A), which were used to benchmark the new method. PartMC-MOSAIC tracks the masses of 20 aerosol components, including water, in thousands of computational particles, corresponding to a stochastic representation of the 20-dimensional distribution that describes aerosol mixing state. The procedure for computing CCN activity from a particle-resolved population is shown in Figure 1a–1c. The particle-resolved simulations reveal complex variation in particle size and composition, as illustrated by the number density distribution with respect to two coordinates: dry diameter  $D_{\text{dry}}$ , which is the volume-equivalent diameter for all species except water, and effective hygroscopicity parameter  $\kappa$ , which is the volume-weighted hygroscopicity parameter for each constituent aerosol species [Petters and Kreidenweis, 2007]. The bivariate distribution  $n(D_{\text{dry}}, \kappa)$  is the reduced form of the 20-dimensional distribution with respect to aerosol components that is most relevant for CCN activation, but the procedure described in this study can be extended to different projections of the particle size-composition distribution for aerosol dynamics simulations and calculation of aerosol optical properties. The bivariate number distribution  $n(D_{\text{dry}}, \kappa)$  for the benchmark population is shown in Figure 1a, where  $n(D_{\text{dry}}, \kappa)$  is normalized by the total particle number concentration. Because the particle-resolved model resolves composition at the particle level, it is uniquely suited for benchmarking approximate aerosol representations, including the quadrature-based approach discussed here.

For the particle-resolved population, the normalized number distribution with respect to critical supersaturation,  $n(s_c)$ , is shown in Figure 1b. The critical supersaturation at which each simulated particle becomes CCN-active,  $s_c$ , is computed as a function of  $D_{\text{dry}}$  and  $\kappa$  (see Appendix B). The overall number fraction of particles able to serve as CCN at each supersaturation threshold  $N_{\text{CCN}}(s)/N$ , where  $N$  is the total particle number concentration, is then computed as the cumulative distribution of  $n(s_c)$ , shown for this benchmark population in Figure 1c.

## 2.2 Constructing quadratures from moments of $n(D_{\text{dry}}, \kappa)$

Figure 1d shows the location of quadrature abscissas for a specific quadrature approximation for the bivariate distribution shown in Figure 1a, derived from a specific set of moment constraints  $\mu_k$ . A generalized moment  $\mu_k$  of the distribution  $n(D_{\text{dry}}, \kappa)$  is an integral over  $n(D_{\text{dry}}, \kappa)$  with respect to some kernel function of the coordinates,  $\phi_k(D_{\text{dry}}, \kappa)$ :

$$\mu_k = \int_0^\infty \int_0^\infty n(D_{\text{dry}}, \kappa) \phi_k(D_{\text{dry}}, \kappa) dD_{\text{dry}} d\kappa. \quad (1)$$

For example, for  $\phi_k(D_{\text{dry}}, \kappa) = D_{\text{dry}}^m$ ,  $\mu_k$  is the  $m^{\text{th}}$  power moment with respect to diameter and, therefore, independent of  $\kappa$ . Here, we computed constraints in terms of modified moments [Press *et al.*, 1990], which are linear combinations of the power moments (see Appendix C). Continuity equations, expressed in terms of moments, can be solved exactly in closed form only in very special cases, but approximate closure can be achieved by representing integrals over the number distribution using numerical quadrature. Remarkably, to be accurate, quadrature-based models require tracking only a small number of quadrature abscissas and weights; the distribution itself is not required. The quadrature approximation of  $n(D_{\text{dry}}, \kappa)$  takes the form:

$$\mu_k \approx \sum_{i=1}^N w_i \phi_k(D_{\text{dry},i}, \kappa_i), \quad (2)$$

such that  $\mu_k$ , expressed in terms of the bivariate abscissa coordinates,  $D_{\text{dry},i}$  and  $\kappa_i$ , and weights,  $w_i$ , is identical to the result of Equation 1 for  $k = 1, \dots, N_q$ . In this way, for each  $\mu_k$ , the summation over  $N_q$  quadrature points in Equation 2 is exact. Algorithms have been developed to construct quadrature approximations that satisfy power moments and modified moments of univariate distributions [Gordon, 1968; Press *et al.*, 1990], and these approximations yield on the order of  $N_q/2$  quadrature points for  $N_q$  moment constraints. This paper introduces a new technique for generating a quadrature representation from sets of generalized constraints, which provides a distributed set of  $N_q$  quadrature points for  $N_q$  generalized moments and is also suitable for constructing quadrature for multivariate distributions.

We seek a quadrature approximation that can be derived from a set of known constraints  $\mu_k$ . Ideally, the quadrature abscissas  $x_i$  will be distributed across the variable space in order to resolve key features of the probability distribution. We found that a distributed set of abscissas,  $\{D_{\text{dry},i}, \kappa_i\}$ , and weights,  $w_i$ , can be constructed through a linear program that maximizes an entropy-inspired cost function (see Appendices D and E).

## 2.3 Reconstructing $n(s_c)$ from quadrature

For aerosol representations that track only quadrature points (Figure 1b), a procedure is required if one seeks a corresponding continuous distribution, and often this can be offline. We propose to transform each quadrature point from the bivariate abscissas,  $\{D_{\text{dry},i}, \kappa_i\}$ , into univariate abscissas,  $s_{c,i}$ , for each quadrature point  $i = 1, \dots, N_q$  and preserving the weights, as shown in Figure 1c. The moments of  $n(s_c)$ ,  $\mu_l$  can then be estimated using the transformed quadrature:

$$\mu_l = \sum_{i=1}^N w_i \phi_l(s_{c,i}). \quad (3)$$

The number of moments in  $s_c$  space,  $N_s$ , need not be the same as the number of moments in  $D_{\text{dry}}\text{-}\kappa$  space,  $N_q$ .

The desired continuous distribution  $n(s_c)$ , shown in Figure 1f, is constructed by finding the distribution having maximum entropy, given the approximated moments in  $s_c$  space,  $\hat{\mu}_l$  for  $l = 1, \dots, N_s$  (see Appendix D). We use  $\hat{\mu}_l$  to indicate the moments in  $s_c$  space approximated from the quadrature and  $\mu_l$  to indicate exact moments computed from the benchmark population, which would not be known in a quadrature-based simulation. Similarly, we denote the reconstructed distribution as  $\hat{n}(s_c)$  to

differentiate between the benchmark distribution  $n(s_c)$ . If only the moments of a distribution are known, the distribution that best represents the state of knowledge afforded by the moment constraints is the distribution having maximal entropy [Jaynes, 1957a,b]. Distributions of maximum entropy can be constructed for any set of moment constraints and are not limited to univariate aerosol distributions, but here we confine ourselves to the univariate case. In Section 3, we show that the accuracy of the reconstructed distribution depends strongly on the set of moments  $\mu_l$  chosen to reconstruct the distribution and the accuracy with which the moments approximated from the quadrature  $\hat{\mu}_l$  represent the moments taken directly from the particle-resolved distribution  $\mu_l$ .

## 2.4 Evaluating error in $N_{\text{CCN}}(s)/N$

The aim of this study is to find a moment-based representation that accurately reproduces CCN activation properties of benchmark aerosol populations from the particle-resolved model PartMC-MOSAIC. To this end, we apply a genetic algorithm (see Appendix F) to find optimal sets of bivariate moments to be constrained in constructing the bivariate quadrature. We define  $N_{\text{CCN}}(s)$  and  $\hat{N}_{\text{CCN}}(s)$  as the number concentration of CCN from the particle-resolved model and from a quadrature representation, respectively, and  $N$  is the overall number concentration of particles, which is constrained to be identical in both cases. The error  $\varepsilon$  is quantified as the mean squared error:

$$\varepsilon = \frac{1}{N} \int_0^\infty (\hat{N}_{\text{CCN}}(s) - N_{\text{CCN}}(s))^2 ds. \quad (4)$$

The following sections describe methods for identifying optimized sets of moment constraints for constructing the quadrature in  $D_{\text{dry}}-\kappa$  space  $\mu_k$  and optimized sets of moments for reconstructing continuous distributions in  $s_c$  space. We define optimized moment sets as those yielding the best representation of the CCN activation spectrum, indicated by small mean squared error  $\varepsilon$ .

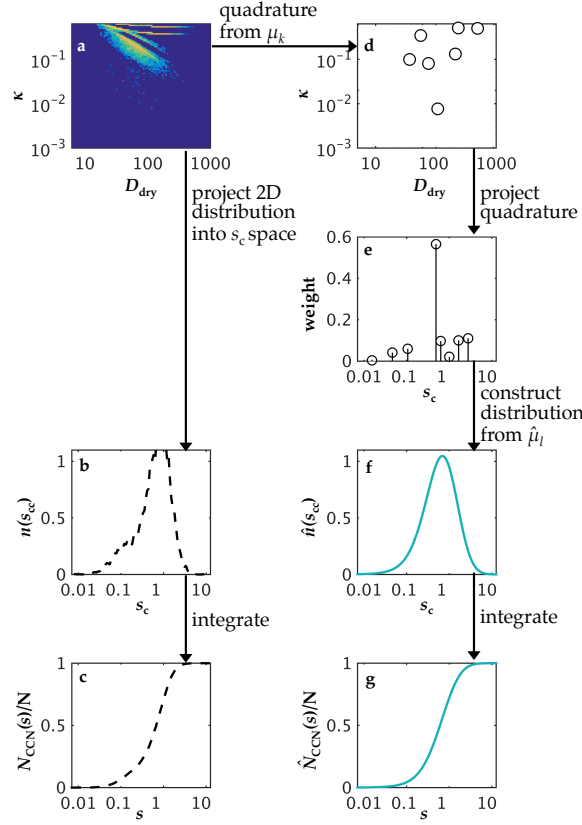
## 3 Moment constraints for aerosol simulations

Although the proposed maximum-entropy approaches for constructing quadrature approximations and continuous distributions can yield accurate representations of the benchmark populations, here we show that this accuracy depends strongly on the selection of moment constraints. The moments with respect to  $s_c$  computed from the bivariate quadrature,  $\hat{\mu}_l$  are used to construct the continuous distribution  $\hat{n}(s_c)$  (see Section 2.3); Section 3.1 introduces the necessary and sufficient set of constraints  $\mu_l$  to accurately reconstruct  $n(s_c)$ . The accuracy of the continuous distribution  $\hat{n}(s_c)$  depends on the degree to which the projected quadrature points, which are originally constructed in  $D_{\text{dry}}-\kappa$  space (see Section 2.2), approximate the true moments of  $n(s_c)$  for the benchmark distribution; Section 3.2 describes a procedure for selecting moment constraints of  $n(D_{\text{dry}}, \kappa)$  required to construct accurate quadrature approximations of  $n(D_{\text{dry}}, \kappa)$  and  $\hat{n}(s_c)$ .

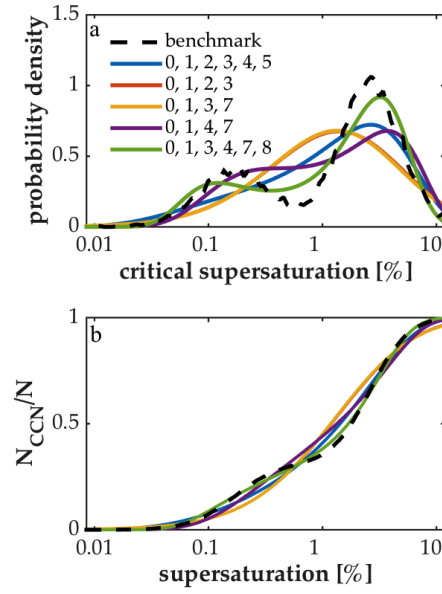
### 3.1 Moments of $n(s_c)$ for reconstructing distributions

In this section, we identify the set of moments,  $\mu_l$  for  $l = 1, \dots, N_s$ , that yield highest accuracy for the reconstructed distribution  $\hat{n}(s_c)$ . We show that the accuracy of the reconstruction depends on the degree to which the quadrature is able to reproduce the moments taken directly from the particle-resolved distribution  $n(s_c)$ .

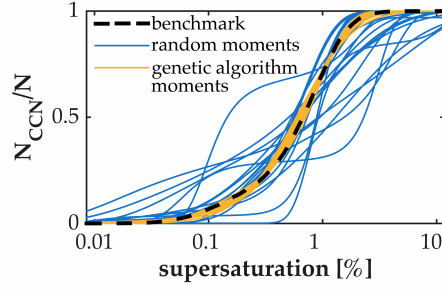
Figure 2a shows distributions constructed from the quadrature using selected combinations of moments  $\mu_l$ . We found that  $n(s_c)$  can be estimated with high accuracy using the six modified moments  $\mu_l$ , corresponding to Legendre polynomials of power  $l = 0, 1, 3, 4, 7, 8$ , computed with respect to  $\ln s_c$  (see Appendix C). Provided these six moments can be computed with high accuracy from projections of the quadrature approximation of  $n(D_{\text{dry}}, \kappa)$ , the moment-based representation reproduces the CCN spectrum  $N_{\text{CCN}}(s)/N$  of the benchmark populations with high accuracy, as shown through comparison between green and black lines in Figure 2b.



**Figure 1.** Procedure for computing CCN spectrum from particle-resolved model (a–c) and from quadrature-based representation (d–g): (a) bivariate number distribution with respect to  $D_{\text{dry}}$  and  $\kappa$  from particle-resolved model, (b) number distribution with respect to  $s_c$  from particle-resolved model, (c)  $N_{\text{CCN}}(s)/N$  from particle-resolved model, (d) location of abscissas in  $D_{\text{dry}}-\kappa$  space for quadrature, (e) abscissas transformed into  $s_c$  and weights for quadrature, (f) maximum entropy reconstruction of number density with respect to  $s_c$  from quadrature, and (g)  $N_{\text{CCN}}(s)/N$  for quadrature.



**Figure 2.** Comparison between benchmark (black dashed lines) and reconstructions (colored lines) of (a) number density with respect to  $s_c$  and (b)  $N_{CCN}(s)/N$  for different combinations of moments of  $n(s_c)$ , computed directly from the benchmark population, reveals that only six moments  $\mu_i$  are needed, but these moments must be carefully selected.



**Figure 3.** Comparison between benchmark (black dashed line), reconstructions from randomly selected moments (blue lines), and reconstructions from moments chosen with the genetic algorithm (orange lines) show that care must be taken in selection of of bivariate moments constraints for construction of the quadrature.

### 3.2 Moments of $n(D_{\text{dry}}, \kappa)$ for constructing bivariate quadrature

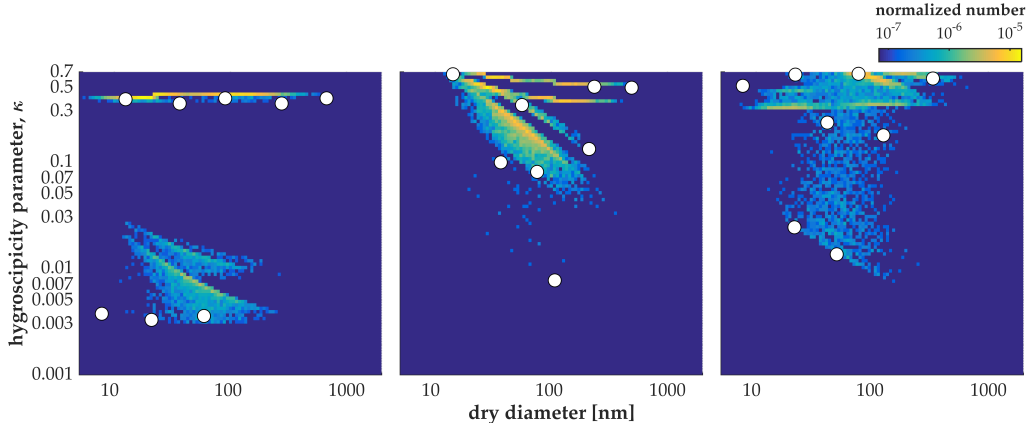
A good bivariate quadrature representation of the aerosol must reproduce key moments of  $n(s_c)$  on projection, which are used to construct the CCN activation spectrum  $N_{\text{CCN}}(s)/N$ . Selecting the optimal six moments from ten possible combinations, as was done for selecting key moments of  $n(s_c)$ , requires testing a total of 210 combinations. On the other hand, selecting optimal combinations of bivariate moments requires testing unfeasibly large sets of combinations. For example, here we identify combinations of bivariate Legendre moments of power  $n = -3, \dots, 8$  (Equation 8) for each of the two variables,  $D_{\text{dry}}$  and  $\kappa$ , yielding 144 possible bivariate moments. Selecting the eight optimal moments from the 144 candidates would require performing the full procedure for  $3.8 \times 10^{11}$  combinations. Instead, we applied a genetic algorithm to find suitable sets of bivariate moments  $\mu_k$ ,  $k = 1, \dots, N_q$ , defined by Equations 1 and 2, that should be tracked in moment-based models for accurate representation of CCN properties.

For a single population, the CCN spectra computed from the quadrature,  $\hat{N}_{\text{CCN}}(s)/N$ , is compared with the particle-resolved CCN spectrum,  $N_{\text{CCN}}(s)/N$  in Figure 3, using 20 randomly selected combinations (blue lines) of  $N_q = 8$  moments and for 20 combinations that were selected using the genetic algorithm (orange lines). Although the proposed procedure yields high accuracy in approximated CCN spectra  $\hat{N}_{\text{CCN}}(s)/N$  for those moments selected by the genetic algorithm, the large errors in  $\hat{N}_{\text{CCN}}(s)/N$  for randomly chosen moments illustrate that the accuracy of the procedure depends strongly on the choice of moment constraints used in assignment of the bivariate quadrature points.

## 4 Sparse approximation for efficient and accurate aerosol representation

For the selected moments of  $n(D_{\text{dry}}, \kappa)$ ,  $\mu_k$ , and the selected univariate moments for estimating  $n(s_c)$ ,  $\mu_l$ , the quadrature-based representation accurately reproduces CCN activation for the three bivariate benchmark populations shown in Figure 4. The location of quadrature abscissas associated with a set of  $N_q = 8$  moments (white dots) are superimposed on benchmark dis-





**Figure 4.** Bivariate distributions from particle-resolved model (surface plots) and location of abscissas from the quadrature (white dots) for the three populations, which were sampled at (a) 7:00 am, after 1 hour of simulation, (b) 12:00 pm, after 6 hours of simulation, and (c) 6:00 am the following day, after 24 hours of simulation. The corresponding weights are not shown here but are shown by the white dots in the one-dimensional projection plots of Figure 5. In all cases, the maximum-entropy-inspired technique yields a distributed set of abscissas.

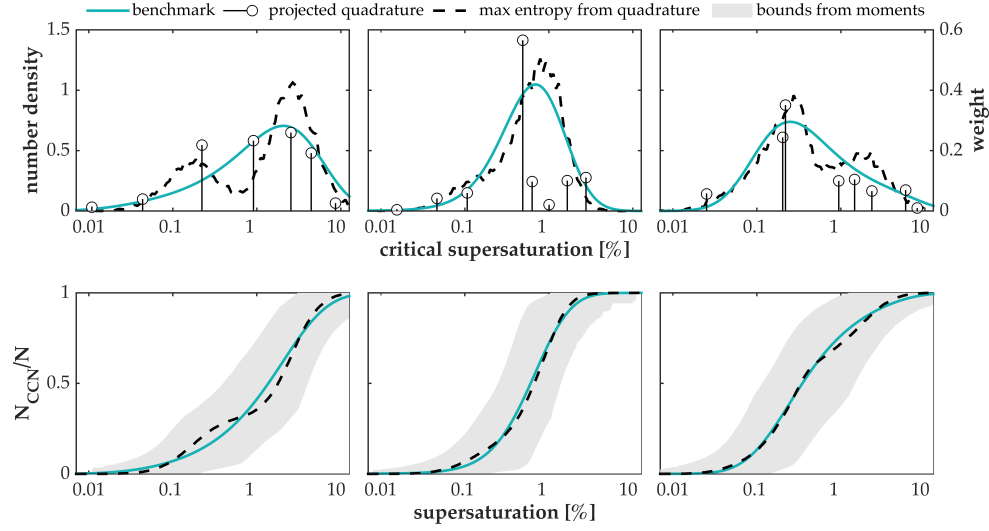
tributions (surface plots). The quadrature was constructed using the approach outlined in Section 2.2 for moments of  $n(D_{\text{dry}}, \kappa)$ , where the specific combinations of moment constraints  $\mu_k$  for  $k = 1, \dots, N_q$  were identified using the genetic algorithm (Section 3.2 and Appendix F). Optimized bivariate moments identified with the genetic algorithm are listed in Table 1 of Appendix F. The quadrature represents key features of each benchmark distribution using only moments.

The critical supersaturation  $s_{c,i}$  is computed for each quadrature abscissa  $\{D_{\text{dry},i}, \kappa_i\}$ , shown along with corresponding weights by the stems in Figures 5a–5c. These quadrature points are used to compute the projected moments in  $s_c$  space (Section 3.1), and the full distribution  $\hat{n}_c$  is then reconstructed from the estimated moments (Section 2.3). The reconstructed distributions (blue lines in Figures 5a–5c) represent key features of the benchmark distributions (black lines). Although the distributions  $\hat{n}(s_c)$  are not reconstructed exactly, the CCN activation spectra,  $\hat{N}_{\text{CCN}}/N$ , that is computed by integrating over  $\hat{n}(s_c)$  reproduces the benchmark spectrum  $N_{\text{CCN}}(s)/N$  with high accuracy (Figures 5d–5f).

The shaded region of Figures 5d–5f show the bounds on possible solutions for each CCN spectrum, given the modified moments  $\mu_k$  that were used to construct the quadrature approximations in Figure 4; all possible distributions that satisfy the moment constraints  $\mu_k$  lie within these extreme bounds. The large bounds in the shaded region indicate that constraints on  $n(D_{\text{dry}}, \kappa)$  do not, on their own, sufficiently constrain CCN activation properties. On the other hand, adding the maximum-entropy-inspired approach described here accurately represents CCN activation spectra across various particle-resolved distributions.

## 5 Discussion and Conclusions

This manuscript introduces a new procedure for inverting between moments, quadratures, and full distributions, which can be used to advance quadrature-based simulations of aerosol dynamics. By applying a linear program, with a maximum-entropy-inspired cost function, we generated sets of distributed quadrature points that are an ideal model representation of atmospheric aerosols, where a genetic algorithm was used to identify optimal combinations of modified bivariate moments to



**Figure 5.** (a–c) The number distribution with respect to  $s_c$  for the benchmark (dashed curves) are compared with reconstructions from the quadrature (blue curves), where the projection of the underlying quadrature abscissas and weights, indicated by the location and magnitude of the vertical stems. (d–f) The bivariate moments used to construct the quadrature do not effectively constrain  $N_{CCN}(s)/N$ , where bounds on possible solutions are shown by the shading, but the continuous distributions reconstructed from the quadrature (blue curves) accurately represent the benchmark (black dashed curves). Results are shown for the three particle-resolved populations from Figure 4, which were sampled at (a,d) 7:00 am, after 1 hour of simulation, (b,e) 12:00 pm, after 6 hours of simulation, and (c,f) 6:00 am the following day, after 24 hours of simulation.

be used as constraints in the linear program. We found that CCN activation spectra can be computed with high accuracy by reconstructing univariate distributions with respect to  $s_c$ , where moment constraints of  $n(s_c)$  were determined using a variable transform from bivariate quadrature points  $\{D_{\text{dry},i}, \kappa_i\}$  to univariate points  $s_{c,i}$ . The combined procedure yields high accuracy in CCN spectra in comparison with particle-resolved benchmark populations. Although the present manuscript illustrates the efficacy of the proposed procedure for estimating the CCN activation spectra, which is notoriously difficult for moment-based aerosol representations [Wright *et al.*, 2001, 2002], the approach is not limited to the variable spaces that we have chosen to analyze here. In a future study, we will demonstrate how the procedure can be applied to advance quadrature-based aerosol simulations and for computing other quantities of interest, such as aerosol optical properties.

In this study, the traditional quadrature method of moments has been augmented with a maximum-entropy-inspired linear program for quadrature generation and with maximum-entropy spectral analysis to obtain continuous representations of consistent with specified moment constraints. We show that the full particle-resolved data set of over 10,000 particles can be reduced to a sparse set of just eight weighted particles to achieve accurate recovery of CCN activation properties. We show that the quadrature method of moments, combined with maximum entropy methods, can be used to overcome the limitations of traditional moment methods.

## Appendices

### A Benchmark Populations from Particle-Resolved Model

PartMC-MOSAIC simulates the evolution of trace gases and aerosol particles in a well-mixed air parcel. The model tracks the mass of  $j = 1, \dots, A$  constituent aerosol species in each particle  $i = 1, \dots, N_p$ , where  $A$  is the total number of aerosol species, including water, and  $N_p$  is the total number of particles in the population. In the present implementation  $A = 20$  and  $N_p$  is on the order of  $10^5$  particles. The mass of each component in each particle is given by  $m_{j,i}$ , and the mass composition of each particle is represented by the vector  $\hat{m}_i = [m_{1,i}, \dots, m_{j,i}, \dots, m_{A,i}]$ . The model simulates the aerosol evolution due to condensation and evaporation of semi-volatile gases, coagulation between particles, dilution of background air, and particle emissions. A full description of the model is described in [Riemer *et al.*, 2009], and the bivariate distributions with respect to  $D_{\text{dry}}$  and  $\kappa$  are introduced in Fierce *et al.* [2013].

### B Critical supersaturation for CCN activation

The critical supersaturation ( $s_{c,i}$ ) was computed for each particle  $i$  using the  $\kappa$ -Köhler model, which depends on a particle's overall dry volume, its effective hygroscopicity parameter  $\kappa_i$ , and environmental properties. The equilibrium saturation ratio ( $S_i$ ) over an aqueous droplet is computed through the  $\kappa$ -Köhler model [Petters and Kreidenweis, 2007] as:

$$S_i(D_i) = \frac{D_i^3 - D_{\text{dry},i}^3}{D_i^3 - D_{\text{dry},i}^3(1 - \kappa_i)} \exp\left(\frac{4\sigma_w M_w}{RT\rho_w D_i}\right), \quad (5)$$

where the wet particle volume  $D_i$  is the particle wet diameter,  $T$  is the ambient temperature,  $R$  is the universal gas constant,  $M_w$  is the molecular weight of water, and  $\sigma_w$  is the surface tension of the air-water interface. The effective hygroscopicity parameter  $\kappa_i$  is the volume-weighted average of the hygroscopicity parameter  $\kappa$  of the particle's constituent aerosol species:

$$\kappa_i = \frac{\sum_k^{A-1} v_{k,i} \kappa_k}{\sum_k^{A-1} v_{k,i}}. \quad (6)$$

Values for  $\kappa$  for each species are given in [Fierce *et al.*, 2013].

The critical wet diameter is the diameter at which  $S_i(D_i)$  is maximal, and the critical saturation ratio ( $S_{c,i}$ ) is the saturation ratio corresponding to this critical wet diameter. The critical supersaturation  $s_{c,i}$  is then given by  $s_{c,i} = (S_{c,i} - 1) \times 100$ . The number concentration of CCN at each environmental supersaturation  $s$  is computed as the number concentration of particles that will activate into at that  $s$ , that is the total number concentration of particles per volume having a critical supersaturation  $s_{c,i} \leq s$ .

### C Modified Moments from Legendre Polynomials

This study uses modified moments based on Legendre polynomials, which are better conditioned than traditional geometric moments. The  $l^{\text{th}}$  modified moment over the univariate critical supersaturation distribution  $n(s_c)$  is defined as the integral over the  $l^{\text{th}}$  Legendre polynomial  $\phi_l$  of  $\ln s_c$ :

$$\phi_l(s_c) = \frac{1}{2^l l!} \left[ \frac{d^l}{d(\ln s_c)^l} (\ln^2 s_c - 1)^l \right]. \quad (7)$$

Legendre polynomials are defined for  $l \geq 0$ . We define modified moments of  $n(s_c)$  for  $l < 0$  as integrals over the  $l^{\text{th}}$  Legendre polynomials for  $\ln s_c^{-1}$ . For example, the Legendre polynomials of order 3 and  $-3$  are given by  $\phi_3(s_c) = \frac{1}{2}(5 \ln^3 s_c - 3 \ln s_c)$  and  $\phi_{-3}(s_c) = \frac{1}{2}(5 \ln^{-3} s_c - 3 \ln^{-1} s_c)$ , respectively.

For the bivariate distributions  $n(D_{\text{dry}}, \kappa)$ , each modified moment  $\mu_k$ , is computed as the integral over the kernel function  $\phi_k(D_{\text{dry}}, \kappa)$ . We define the modified kernel function  $\phi_k$  as the product of the  $m^{\text{th}}$  Legendre polynomial of  $\ln D_{\text{dry}}$  and the  $n^{\text{th}}$  Legendre polynomial of  $\ln \kappa$ :

$$\phi_k(D_{\text{dry}}, \kappa) = \left( \frac{1}{2^m m!} \left[ \frac{d^m}{d(\ln D_{\text{dry}})^m} (\ln^2 D_{\text{dry}} - 1)^m \right] \right) \left( \frac{1}{2^n n!} \left[ \frac{d^n}{d(\ln \kappa)^n} (\ln^2 \kappa - 1)^n \right] \right). \quad (8)$$

### D Distributional Entropy

The entropy  $H_s$  of the reconstructed univariate distribution  $\hat{n}(s_c)$ , which is normalized by total particle number concentration, is given by:

$$H_s = \int_0^\infty \hat{n}(s_c) \ln \hat{n}(s_c) d \ln s_c. \quad (9)$$

Jaynes [1957a,b] showed that if  $n(s_c)$  is the density distribution having maximum entropy, given a set of constraints  $\mu_l$  for  $l = 1, \dots, N_s$ ,  $n(s_c)$  will take the form:

$$\hat{n}(s_c) = \exp \left( - \sum_{l=1}^{N_s} \lambda_l \phi_l(s_c) \right), \quad (10)$$

where  $\lambda_l$  is the Lagrange multiplier of the entropy function  $H_s$  for  $\mu_l$ .

Similarly, the entropy  $H_q$  for the bivariate number density distribution  $n(D_{\text{dry}}, \kappa)$  is given by:

$$H_q = \int_{-\infty}^\infty \int_{-\infty}^\infty n(D_{\text{dry}}, \kappa) \ln n(D_{\text{dry}}, \kappa) d \ln D_{\text{dry}} d \ln \kappa. \quad (11)$$

### E Linear Program to Compute Quadrature Approximation and CCN bounds

A linear program was used to construct the optimized quadrature, with abscissa locations shown in Figure 4, and to bound  $N_{\text{CCN}}(s)/N$ , shown by shading in Figure 5. The linear program maximizes some cost function  $c(D_{\text{dry},i}, \kappa_i)$  subject to specified

constraints  $\mu_k$  (Equation 1) and the requirement that  $w_i \geq 0$ :

$$\text{maximize} \quad - \sum_{i=1}^{N_{\text{grid}}} w_i c(D_{\text{dry},i}, \kappa_i) \quad (12)$$

$$\text{subject to} \quad \sum_{i=1}^{N_{\text{grid}}} w_i \phi_k(D_{\text{dry},i}, \kappa_i) = \mu_k, \quad (13)$$

$$w_i \geq 0, \quad i = 1, \dots, N_{\text{grid}}. \quad (14)$$

The application of linear programming to construct numerical quadrature was introduced in *McGraw* [2013], but we now extend this approach to find optimized sets of quadrature points by maximizing an entropy-inspired cost function:

$$c(D_{\text{dry},i}, \kappa_i) = - \sum_k \lambda_k \phi_k(D_{\text{dry},i}, \kappa_i). \quad (15)$$

The bounds on  $N_{\text{CCN}}(s)/N$  are computed using two linear programs at each supersaturation threshold  $s$ , one for the upper bound and one for the lower bound. The maximum value for  $N_{\text{CCN}}(s)/N$  for some specified  $s$  is computed by maximizing the number of particles with  $s_{c,i} \leq s$ , which is computed from the linear program with the following cost function:

$$c(D_{\text{dry},i}, \kappa_i) = s_c(D_{\text{dry},i}, \kappa_i) \leq s. \quad (16)$$

Similarly, the minimum  $N_{\text{CCN}}/N$  is computed using a cost function that maximizes the number of particles with  $s_{c,i} \geq s$ :

$$c(D_{\text{dry},i}, \kappa_i) = s_c(D_{\text{dry},i}, \kappa_i) \geq s. \quad (17)$$

The linear program is applied on a grid  $x_i$  for  $i = 1, \dots, N$ , where  $N$  is much larger than the number of moment constraints. Using the dual simplex algorithm [Lemke, 1954], the linear program yields only a subset of non-zero weights  $w_i$ , where the number of non-zero values for  $w_i$  is equal to the number of moment constraints. It is these non-zero weights  $w_i$  and associated abscissas  $x_i$  that comprise the sparse set of quadrature points.

## F Genetic Algorithm to Optimize Constraints

We applied a genetic algorithm to find optimal sets of moments for constructing the bivariate quadrature. Genetic algorithms are a class of optimization methods inspired by natural selection, first introduced by *Holland* [1975]. Implementation of a genetic algorithm requires a genetic encoding of possible solutions, in this case as a series of 0's and 1's, and fitness function to be optimized, in this case minimization of error in  $N_{\text{CCN}}(s)/N$  across distinct populations. The algorithm is initiated with a population of candidate solutions. In this study, the initial set of candidate populations was chosen using Latin hypercube sampling. The optimal order of Legendre polynomials for computing the modified moments with respect to  $\ln D_{\text{dry}}$  and  $\ln \kappa$  (Equation 8) identified with the genetic algorithm are outlined in Table 1.

## Acknowledgements

LMF is supported by the University Corporation for Atmospheric Research under a NOAA Climate & Global Global Change Postdoctoral Fellowship. RLM and LMF acknowledge support by the Atmospheric Systems Research Program of the US Department of Energy. The model (PartMC Version 2.1.6) and input files for the particle-resolved model simulations are available at <http://lagrange.mechse.illinois.edu/partmc/>. All other methods supporting the conclusions are described in Section 2 and in the Appendices.

**Table 1.** Kernel function  $\phi_k$  used to generate optimized bivariate moments  $\mu_k = \sum_i^{N_q} \phi_k$ , which were identified using the genetic algorithm.

---


$$\begin{aligned}
\phi_1 &= \frac{1}{8} (35 \ln^4 D_{\text{dry}} - 30 \ln^2 D_{\text{dry}} + 3) \\
\phi_2 &= \ln^{-1} \kappa \\
\phi_3 &= \frac{1}{2} (3 \ln^{-2} \kappa - 1) \\
\phi_4 &= \frac{1}{2} (3 \ln^2 D_{\text{dry}} - 1) \ln \kappa \\
\phi_5 &= \ln D_{\text{dry}} \frac{1}{2} (3 \ln^{-2} \kappa - 1) \\
\phi_6 &= \frac{1}{2} (3 \ln^{-2} D_{\text{dry}} - 1) \frac{1}{2} (3 \ln^{-2} \kappa - 1) \\
\phi_7 &= \frac{1}{2} (3 \ln^{-2} D_{\text{dry}} - 1) \frac{1}{8} (35 \ln^4 \kappa - 30 \ln^2 \kappa + 3) \\
\phi_8 &= \frac{1}{8} (63 \ln^5 D_{\text{dry}} - 70 \ln^3 D_{\text{dry}} + 15 \ln D_{\text{dry}}) \frac{1}{8} (63 \ln^5 \kappa - 70 \ln^3 \kappa + 15 \kappa)
\end{aligned}$$


---

## References

- Bauer, S. E., D. L. Wright, D. Koch, E. R. Lewis, R. McGraw, L. S. Chang, S. E. Schwartz, and R. Ruedy (2008), MATRIX (multiconfiguration Aerosol TRacker of mIXing state): an aerosol microphysical module for global atmospheric models, *Atmospheric Chemistry and Physics*, 8(20), 6003–6035.
- Binkowski, F. S., and S. J. Roselle (2003), Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component: 1. Model description, *Journal of Geophysical Research*, 108(4183), 15.
- Carslaw, K., L. Lee, C. Reddington, K. Pringle, A. Rap, P. Forster, G. Mann, D. Spracklen, M. Woodhouse, L. Regayre, et al. (2013), Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503(7474), 67–71.
- Chung, C. E., K. Lee, and D. Müller (2011), Effect of internal mixture on black carbon radiative forcing, *Tellus B*, 64.
- Cooke, W. F., C. Liou, H. Cachier, and J. Feichter (1999), Construction of a  $1 \times 1$  fossil fuel emission data set for carbonaceous aerosol and implementation and radiative impact in the ECHAM4 model, *Journal of Geophysical Research*, 104(D18), 22,137–22,162.
- Cubison, M., B. Ervens, G. Feingold, K. Docherty, I. Ulbrich, L. Shields, K. Prather, S. Hering, and J. Jimenez (2008), The influence of chemical composition and mixing state of Los Angeles urban aerosol on CCN number and cloud properties, *Atmospheric Chemistry and Physics*, 8(18), 5649–5667.
- Ervens, B., M. J. Cubison, E. Andrews, G. Feingold, J. A. Ogren, J. L. Jimenez, P. K. Quinn, T. S. Bates, J. Wang, Q. Zhang, H. Coe, M. Flynn, and J. D. Allan (2010), CCN predictions using simplified assumptions of organic aerosol composition and mixing state: a synthesis from six different locations, *Atmospheric Chemistry and Physics*, 10, 4795–4807.
- Fierce, L., N. Riemer, and T. C. Bond (2013), When is cloud condensation nuclei activity sensitive to particle characteristics at emission?, *Journal of Geophysical Research*, 118(24).
- Fierce, L., T. C. Bond, S. E. Bauer, F. Mena, and N. Riemer (2016a), Black carbon absorption at the global scale is affected by particle-scale diversity in composition, *Nature Communications*, 7.
- Fierce, L., N. Riemer, and T. Bond (2016b), Toward reduced representation of mixing state for simulating aerosol effects on climate, *Bulletin of the American Meteorological Society*, 0(0), Early Online Release, doi:10.1175/BAMS-D-16-0028.1.
- Gordon, R. G. (1968), Error bounds in equilibrium statistical mechanics, *Journal of Mathematical Physics*, 9(5), 655–663.
- Haywood, J. M., D. L. Roberts, A. Slingo, J. M. Edwards, and K. P. Shine (1997), General circulation model calculations of the direct radiative forcing by anthropogenic sulfate and fossil-fuel soot aerosol, *Journal of Climate*, 10(7), 1562–1577.
- Holland, J. H. (1975), *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.*, U Michigan Press.

- Hulburt, H. M., and S. Katz (1964), Some problems in particle technology: A statistical mechanical formulation, *Chemical Engineering Science*, 19(8), 555–574.
- Jacobson, M. (1997), Development and application of a new air pollution modeling system–II. Aerosol module structure and design, *Atmospheric Environment*, 31(2), 131–144.
- Jacobson, M. Z. (2001), Strong radiative heating due to the mixing state of black carbon in atmospheric aerosols, *Nature*, 409(6821), 695–697.
- Jacobson, M. Z. (2002), Analysis of aerosol interactions with numerical techniques for solving coagulation, nucleation, condensation, dissolution, and reversible chemistry among multiple size distributions, *Journal of Geophysical Research*, 107(D19), AAC–2.
- Jaynes, E. T. (1957a), Information theory and statistical mechanics, *Physical review*, 106(4), 620.
- Jaynes, E. T. (1957b), Information theory and statistical mechanics. ii, *Physical review*, 108(2), 171.
- Kaufman, Y. J., D. Tanré, and O. Boucher (2002), A satellite view of aerosols in the climate system, *Nature*, 419(6903), 215–223.
- Kleeman, M. J., and G. R. Cass (2001), A 3d Eulerian source-oriented model for an externally mixed aerosol, *Environmental science & technology*, 35(24), 4834–4848.
- Krieger, U. K., C. Marcolli, and J. P. Reid (2012), Exploring the complexity of aerosol particle properties and processes using single particle techniques, *Chemical Society Reviews*, 41(19), 6631–6662.
- Lemke, C. E. (1954), The dual method of solving the linear programming problem, *Naval Research Logistics Quarterly*, 1(1), 36–47.
- Lesins, G., P. Chylek, and U. Lohmann (2002), A study of internal and external mixing scenarios and its effect on aerosol optical properties and direct radiative forcing, *Journal of Geophysical Research*, 107(D10), AAC–5.
- Mann, G., K. Carslaw, D. Ridley, D. Spracklen, K. Pringle, J. Merikanto, H. Korhonen, J. Schwarz, L. Lee, P. Manktelow, et al. (2012), Intercomparison of modal and sectional aerosol microphysics representations within the same 3-d global chemical transport model, *Atmospheric Chemistry and Physics*, 12(10), 4449–4476.
- Matsui, H., M. Koike, Y. Kondo, N. Moteki, J. D. Fast, and R. A. Zaveri (2013), Development and validation of a black carbon mixing state resolved three-dimensional model: Aging processes and radiative impact, *Journal of Geophysical Research*, 118(5), 2304–2326.
- McGraw, R. (1997), Description of aerosol dynamics by the quadrature method of moments, *Aerosol Science and Technology*, 27(2), 255–265.
- McGraw, R. (2013), Sparse aerosol models beyond the quadrature method of moments, in *AIP Conf. Proc.*, vol. 1527, pp. 651–654.
- Myhre, G., F. Stordal, K. Restad, and I. S. Isaksen (1998), Estimation of the direct radiative forcing due to sulfate and soot aerosols, *Tellus B*, 50(5).
- Myhre, G., A. Myhre, and F. Stordal (2001), Historical evolution of radiative forcing of climate, *Atmospheric Environment*, 35(13), 2361–2373.
- Oshima, N., M. Koike, Y. Zhang, Y. Kondo, N. Moteki, N. Takegawa, and Y. Miyazaki (2009), Aging of black carbon in outflow from anthropogenic sources using a mixing state resolved model: Model development and evaluation, *Journal of Geophysical Research*, 114(D6).
- Penner, J. E., C. C. Chuang, and K. Grant (1998), Climate forcing by carbonaceous and sulfate aerosols, *Climate Dynamics*, 14(12), 839–851.
- Petters, M. D., and S. M. Kreidenweis (2007), A single parameter representation of hygroscopic growth and cloud condensation nucleus activity, *Atmospheric Chemistry and Physics*, 7(8), 1961–1971.
- Pirjola, L., S. Tsyro, L. Tarrason, and M. Kulmala (2003), A monodisperse aerosol dynamics module, a promising candidate for use in long-range transport models: Box model tests, *Journal of Geophysical Research*, 108(D9).
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling (1990), Numerical recipes.

- Riemer, N., M. West, R. A. Zaveri, and R. C. Easter (2009), Simulating the evolution of soot mixing state with a particle-resolved aerosol model, *Journal of Geophysical Research*, *114*(D9), D09,202.
- Stier, P., J. Feichter, S. Kinne, S. Kloster, E. Vignati, J. Wilson, L. Ganzeveld, I. Tegen, M. Werner, Y. Balkanski, M. Schulz, O. Boucher, A. Minikin, and A. Petzold (2005), The aerosol-climate model echam5-ham, *Atmospheric Chemistry and Physics*, *5*(4), 1125–1156.
- Stocker, T., D. Qin, G. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, B. Bex, and B. Midgley (2013), Ipcc, 2013: climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change.
- Wexler, A. S., F. W. Lurmann, and J. H. Seinfeld (1994), Modelling urban and regional aerosols—I. model development, *Atmospheric Environment*, *28*(3), 531–546.
- Wright, D., P. Kasibhatla, R. McGraw, and S. Schwartz (2001), Description and evaluation of a six-moment aerosol microphysical module for use in atmospheric chemical transport models, *Journal of Geophysical Research*, *106*(D17), 20,275.
- Wright, D., S. Yu, P. Kasibhatla, R. McGraw, S. Schwartz, V. Saxena, and G. Yue (2002), Retrieval of aerosol properties from moments of the particle size distribution for kernels involving the step function: Cloud droplet activation, *Journal of Aerosol Science*, *33*(2), 319–337.
- Zaveri, R., R. Easter, J. Fast, and L. Peters (2008), Model for simulating aerosol interactions and chemistry (MOSAIC), *Journal of Geophysical Research*, *113*, D13,204.
- Zhang, G., X. Bi, J. He, D. Chen, L. Y. Chan, G. Xie, X. Wang, G. Sheng, J. Fu, and Z. Zhou (2014), Variation of secondary coatings associated with elemental carbon by single particle analysis, *Atmospheric Environment*, *92*, 162–170.
- Zhu, S., K. Sartelet, and C. Seigneur (2015), A size-composition resolved aerosol model for simulating the dynamics of externally mixed particles: Scram (v 1.0), *Geoscientific Model Development*, *8*(6), 1595–1612.